MM Adapted MH Methods

Yosra Marnissi, Emilie Chouzenoux, Amel Benazza-Benyahia, Jean-Christophe Pesquet

MM Adapted MH Methods

Yosra Marnissi*, Emilie Chouzenoux[†§], Amel Benazza-Benyahia[‡] and Jean-Christophe Pesquet[§]
[*] Safran TECH, Groupe Safran 1, Rue Jeunes Bois, 78772 Châteaufort, France.
[§] Center for Visual Computing, CentraleSupélec, INRIA and University Paris Saclay, 3 rue Joliot Curie, 91190 Gif sur Yvette, France
[‡] University of Carthage, COSIM Lab., SUP'COM, Cité Technologique des Communications, 2080, Tunisia

***Abstract*—In this paper, we derive a novel MH proposal, inspired from Langevin dynamics, where the drift term is preconditioned by an adaptive matrix constructed through an MM strategy. We propose several variants of low-complexity curvature metrics applicable to large scale problems. The proposed method is shown to exhibit a good performance in an image recovery example.**

I. Introduction

MCMC methods proceed in two main steps. First, a Markov chain is built with a given transition rule so that its stationary states follow the posterior law. Once the Markov chain has reached its stationary distribution, Monte Carlo approximation is used to infer the posterior characteristics. One typical approach to explore the parameter space is the Random Walk (RW) whose adaptive proposal law takes the form of a Gaussian distribution centered at the current state. The popularity of this algorithm is mainly related to its simplicity of implementation. However, the RW usually takes too many steps to reach stability for high dimensional models. In this respect, a large amount of works has been devoted to construct proposals in Metropolis-Hastings (MH) algorithms having a better performance [1]. In this work, we are interested in proposals based on the Euler discretization of the Langevin stochastic differential equation where the drift term accounts for the slope and curvature of the target law. Our main contribution is to propose a preconditioned version of the standard MH adapted Langevin algorithm using an adaptive matrix based on a Majorize-Minimize (MM) strategy.

II. Proposed approach

To extend the idea behind MM quadratic optimization strategies to the context of stochastic samplers, the idea is to push the proposal distribution of the MH algorithm at each iteration from the current state to a region with high density value. More precisely, the mean of the proposal density is picked by using an MM search step and then the space around this center is explored according to an MM curvature matrix $\mathbf{Q}(\mathbf{x}^{(t)})$ that describes the local variations of the target distribution. This results in a preconditioned Langevin proposal where the scale matrix, equal to the inverse of the curvature matrix $\mathbf{Q}(\mathbf{x}^{(t)})$, is constructed according to the MM strategy.

We focus on the case when the minus-log of the target density function $\mathcal{J} = -\log \mathsf{p}(\cdot \mid \mathbf{z})$ can be expressed up to an additive constant as

$$(\forall \mathbf{x} \in \mathbb{R}^Q) \quad \mathcal{J}(\mathbf{x}) = \Phi(\mathbf{H}\mathbf{x} - \mathbf{z}) + \Psi(\mathbf{x}) \tag{1}$$

where $\mathbf{z} \in \mathbb{R}^N$ is the vector of observed data, $\mathbf{H} \neq \mathbf{0}_{N\times Q}$, and

$$\Psi(\mathbf{x}) = \sum_{s=1}^{S} \psi_s(\|\mathbf{V}_s\mathbf{x} - \mathbf{c}_s\|) \tag{2}$$

with $(\forall s \in \{1,\ldots,S\})$ $\mathbf{V}_s \in \mathbb{R}^{P_s\times Q}$, $\mathbf{c}_s \in \mathbb{R}^{P_s}$, and $(\psi_s)_{1\leqslant s\leqslant S}$ is a set of nonnegative continuous functions.

Under technical assumptions, convex quadratic tangent majorants of (1) can be obtained by setting (see [2]):

$$(\forall \mathbf{x} \in \mathbb{R}^Q) \quad \mathbf{Q}_1(\mathbf{x}) = \mu \mathbf{H}^\top\mathbf{H} + \mathbf{V}^\top \mathrm{Diag}\{\boldsymbol{\omega}(\mathbf{x})\}\mathbf{V} + \zeta\, \mathbf{I}_Q \tag{3}$$

where $\mu \in \,]0+\infty[$, $\mathbf{V} = \left[\mathbf{V}_1^\top,\ldots,\mathbf{V}_S^\top\right]^\top$ and $\boldsymbol{\omega}(\mathbf{x}) = (\omega_i(\mathbf{x}))_{1\leq i\leq P}$ is such that, for every $s \in \{1,\ldots,S\}$ and $p \in \{1,\ldots,P_s\}$,

$$\omega_{P_1+P_2+\ldots+P_{s-1}+p}(\mathbf{x}) = \frac{\dot{\psi}_s(\|\mathbf{V}_s\mathbf{x} - \mathbf{c}_s\|)}{\|\mathbf{V}_s\mathbf{x} - \mathbf{c}_s\|}. \tag{4}$$

Hereabove, ζ is a nonnegative constant that can be useful to ensure the invertibility of $\mathbf{Q}_1(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^Q$. Other simpler constant curvature matrix $\mathbf{Q}_2$ and diagonal curvature matrix $\mathbf{Q}_3(\mathbf{x})$ can be derived by majorizing $\mathbf{Q}_1(\mathbf{x})$.

It is worth emphasizing that the geometric ergodicity of the chain generated by the proposed algorithm can be established for the class of super-exponential distributions.

III. Application to multispectral image denoising

We denoise a Hydice hyperspectral dataset containing images of size $R = 256 \times 256$ over $B = 60$ spectral channels. A multivariate Gaussian likelihood is adopted, whereas a generalized multivariate exponential power distribution is chosen for modelling the prior distribution of the wavelet coefficients of the sought images. We compare the performance of the Gibbs sampler when the posterior law of the wavelet coefficients is explored using various algorithms. Fig. 1 illustrates the evolution of the scale parameter of the prior in the horizontal subband at the first decomposition level.

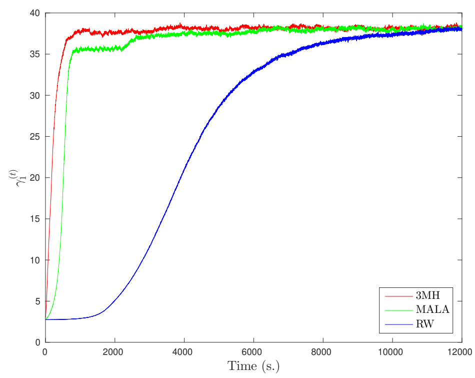

Fig. 1. Convergence speed of RW, MALA and proposed 3MH method.

References

[1] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, Mar. 2016.

[2] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 563–591, 2013.